AD-A278 089

:UMENTATION PAGE

Form Approved
OMB No. 0704-0188

:ed to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and lection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including ters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA work Reduction Project (0704-0188), Washington, DC 20503.

2. REPORT DATE
March 1994

3. REPORT TYPE AND DATES COVERED
Professional Paper

4. TITLE AND SUBTITLE
TACTICAL DECISION MAKING UNDER STRESS: PRELIMINARY RESULTS AND LESSONS LEARNED

5. FUNDING NUMBERS
In-house Funding

6. AUTHOR(S)
S. G. Hutchins and J. T. Kowalski

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Command, Control and Ocean Surveillance Center (NCCOSC)
RDT&E Division
San Diego, CA 92152-5001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Command, Control and Ocean Surveillance Center (NCCOSC)
RDT&E Division
San Diego, CA 92152-5001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. ABSTRACT (Maximum 200 words)

This paper presents preliminary results of ongoing research into tactical decision making under stress (TADMUS). A description will be given of (1) the general methodological approach; (2) the development of the performance measures and issues related to their development; (3) lessons learned in the planning and conducting of this research; and (4) types of errors typically made and their implications for the development of a decision support system (DSS) using a "naturalistic" model of decisionmaking.

Published in *Symposium on Command and Control Research, Basic Research Group, Joint Directors of Laboratories*, July 1993.

14. SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE (Include Area Code)	21c. OFFICE SYMBOL
S. G. Hutchins	(619) 553-9227	Code 442

Tactical Decision Making Under Stress: Preliminary Results and Lessons Learned

Susan G. Hutchins
Information Technology and Decision Science Branch
Naval Command, Control, Communications, and Ocean Surveillance Center
Research, Development, Test and Evaluation Division (NRaD) (Code 442)
53355 Ryne Road Rm 222
San Diego, CA 92152-7252

Jens T. Kowalski
Federal Armed Forces Technical Center for Ships and Naval Weapons
Federal Office of Defense Technology and Procurement
Eckenfoerde, Germany

Abstract

This paper presents preliminary results of ongoing research into tactical decision making under stress (TADMUS). A description will be given of (1) the general methodological approach; (2) the development of the performance measures and issues related to their development; (3) lessons learned in the planning and conducting of this research; and (4) types of errors typically made and their implications for the development of a decision support system (DSS) using a "naturalistic" model of decisionmaking.

Data from ten teams responses to scenarios run in the Decision-Making Evaluation Facility for Tactical Teams (DEFTT) Laboratory will be discussed. Discussion of these results will include a description of the TapRoot Incident Investigation System, the approach used to analyze the data to identify errors. Ways in which these errors can be mitigated by the DSS will also be discussed.

1.0 INTRODUCTION

The Tactical Decision Making Under Stress (TADMUS) program, sponsored by the Office of Naval Research, is being conducted to explore recent developments in decision theory and human-computer interaction technology. Our objective is to apply new decisionmaking models to the design of a decision support system (DSS) for enhancing tactical decisionmaking under highly complex conditions. Data has been collected in the Decision-Making Evaluation Facility for Tactical Teams (DEFTT) Laboratory, a six-station test-bed environment that simulates computer workstations of a Navy Aegis cruiser combat information center (CIC).

DEFTT integrates hardware stations and software modules. The major hardware stations are a Hewlett-Packard (HP) 345 workstation (referred to as the experiment control station [ECS]) with an experiment control software module and six IBM-compatible 386 personal computers. The DEFTT components are interconnected through a local area network. Key features of the DEFTT environment include the following: (1) capability to run complex antiair warfare scenarios; (2) multiple, networked stations to enable configuration of a Navy CIC; (3) ease of scenario authoring and simulation control; (4) the ability to simulate input from a variety of sensors and selectively downgrade them. (For a detailed description of the DEFTT Laboratory see Hutchins & Duffy, 1992.) Actions taken by each team member are recorded as they respond to the highly complex decision events in antiair warfare scenarios. The variables manipulated in the scenarios are workload and ambiguity. Data are used to develop a quantified baseline on performance degradation under these conditions.

2.0 BACKGROUND

The context of this study is the CIC of a Navy Aegis cruiser. The emphasis on Aegis combat

3.0 METHODOLOGICAL APPROACH

Ten pilot sessions have been completed. During each session data was collected on three scenarios; each session took about 3 1/2 - 4 hours to complete. The first 1 1/2 hrs were devoted to an orientation to the DEFTT lab, reviewing the scenario materials (rules of engagement (ROE), political/military background, ship's mission statement, and threat summary), and about one hour of training on use of the DEFTT system.

After receiving training on their individual workstations, the team engages in a practice scenario during which time they can interrupt and ask questions. Before each additional scenario there is a scenario prebrief which includes the specific tasking and current situational update for the upcoming scenario. After engaging in the scenarios we ask subjects to complete various questionnaires to obtain information on participants' reactions to the scenarios, the DEFTT laboratory, demographic information, and stress/workload measurements.

3.1 NASA Task Load Index

The NASA Task Load Index (TLX), a multidimensional rating procedure that provides an overall workload score, is administered to the two key decision makers--the TAO and CO--on IBM compatible PCs. This overall workload score is based on a weighted average of ratings on six subscales: (1) Mental Demands; (2) Physical Demands; (3) Temporal Demands; (4) Own Performance; (5) Effort; and (6) Frustration. The first three dimensions relate to the demands imposed on the subject (Mental, Physical, and Temporal Demands) and the last three to the interaction of a subject with the task (Effort, Frustration, and Performance). The NASA TLX responses are combined to produce overall weighted workload scores. This procedure for collecting workload ratings was developed by the Human Performance Group at NASA Ames Research Center during a three year research effort that involved more than 40 laboratory, simulation, and inflight experiments.

3.2 Scenarios

TADMUS scenarios were intentionally developed to be highly ambiguous so that the decisionmaker is often uncertain about the threat represented by the contact or the intent of a threatening contact. Basic requirements for the scenarios were: (1) operation in shallow and confined waters, (2) neutral and hostile countries in close proximity, (3) modern blue/white systems and weapons among neutral, friendly, and hostile nations, and (4) heavy neutral/friendly traffic in the vicinity. These last three features contribute to creating a scenario with many ambiguous situations, i.e., where loyalties may shift rapidly; third world countries possess first world, high technology weapons; and tactics are ill-defined for the situation.

The scenarios are set in the Persian Gulf where several of the nations initially classified as friendly have reason for suddenly changing their loyalties. Thus, the continued applicability of the rules of engagement is uncertain and the intent of an approaching threat from any one nation is uncertain. Also, the wide distribution of blue/white equipment causes uncertainty as to the national origin of a contact (Riffenburgh, 1991). The scenarios were designed to be independent for experimental control with the intent that a decision in one scenario would not influence decisions in later scenarios. Each scenario follows the pattern of being set in a poorly defined situation where one or more threats of uncertain origin and uncertain intent approach either own ship or the ship being protected and do not respond to warnings. The team must decide on a sequence of responses as the situation evolves.

The lack of a definite ending for the scenarios increases realism because many "real world" incidents, experienced by Navy CIC teams while at sea, do not have a conclusive ending. These incidents involve situations where an aircraft will fly close to a U.S. Navy ship with the intent to embarrass, surveil, or harass the ship as opposed to an intent to attack. Thus the end of the scenario is merely the aircraft flies away--which is the way the TADMUS scenarios end.

3.2.1 Scenario Calibration

Workload comprises three categories of contacts: (1) number of background contacts; (2) number of contacts of interest (COI); and (3) number of critical contacts of interest (CCOI). Background contacts consist of all the usual merchant and commercial aircraft and shipping traffic normally found in a given area. Naturally, this would fluctuate according to the time of day, but would generally remain within some average range for a specific area. In the Persian Gulf it is not unusual to have somewhere between 120 -180 background contacts at one time.

For the Aegis system operator the workload for

during an exercise.) Given the acknowledged variance in the way a CIC team can conduct the task of performing situation assessment, this makes our task a difficult one. One difficulty is the team may take various actions at varying times in response to a specific contact of interest, e.g., challenging the potential threat by radio contact or issuing warnings. Communication with higher authority, or reporting to the officer in tactical command, also allows a degree of latitude in terms of the frequency and the criteria for when a report is required. While there may be definite times when a team would be remiss for not having reported to the officer in tactical command, there are no clear cut guidelines for when the CO must inform the officer in tactical command.

In contrast to other studies about decision making in complex situations, all scenarios used in TADMUS are without a definite end. This means there is no right or wrong outcome at the end, which, on the one hand helps to make the scenarios more realistic, but, on the other hand makes developing measures of performance more difficult. Accordingly, the performance measures focus on the processes which lead to a decision vice more traditional measures, e.g., number of missiles fired, number of missiles that hit the target, etc. An example of focusing on the processes leading to a decision, as opposed to more traditional *outcome* measures, would be an assessment of whether the decisionmaker/s made use of all resources available to them. Did the decisionmaker use all information sources which would have enabled him to perform the most accurate situation assessment and determine intent of an approaching aircraft? Was the decisionmaker able to determine which information was critical to resolving ambiguities? Was the decisionmaker able to identify actions to force the opponent to reveal his intent? Was the decisionmaker able to determine whether the situation actually was evolving a particular way? Was the decisionmaker able to predict which preplanned responses/ counter-measures may be effective if the situation evolves in a certain way?

4.1 Data Collection and Analysis

Data collection and analysis involves three levels. The first level consists of tallying and summarizing the errors documented in the Hot Wash Up Reports. Hot Wash Up Reports are similar to the way the Navy assesses their teams during training; a subject matter expert observes the teams' response to a scenario and then provides feedback on the teams performance, either directly to the team, or to some third party.

The source for the next two levels of data analysis is verbatim transcripts of recorded communications between team members as well as all communications between the team and all other off ship "players" in the scenario. This record of a teams' response to the scenario enables us to extract critical data regarding the cognitive processing they used to respond to the various contacts of interest in each scenario. In the CIC environment all behaviors are verbalized as they are being executed as navy personnel are trained to report all actions taken during an event. We have also found that often the team members will "think aloud" (especially the CO-TAO dyad) regarding their processing of the tactical information. This results in a rich source of data on how the decisionmakers process the information and perform situation assessment.

The second level of analysis involves using criteria-based performance measures to compare with data collected during the experiment. For this phase of performance measure development, we plan to focus on a subset of critical behaviors, (out of a total of some 55 behaviors) to develop a set of measures of performance against which we can compare our observed behavior. For example, issuing warnings is a critical behavior that is usually performed and which provides a critical piece of information to the team in attempting to determine the intent of a contact. (These criteria-based performance measures are still in the process of development.)

The third level of analysis is to apply the TapRoot Incident Investigation System (see 4.2.1 below) which is a very detailed form of analysis where all communications for all team members are entered into a flow diagram, and several levels of analysis are performed on the information.

4.2 Error Analysis

A useful approach for examining the quality of tactical decision making has turned out to be the analysis of human errors. Workload and ambiguity are the variables being manipulated to produce highly ambiguous but realistic scenarios for the TADMUS experiments. Our focus on human errors has proven to be a fertile technique for analyzing the quality of tactical decision making and drawing implications for the decision support system development.

HOT WASH UP REPORTS: TALLY OF ERRORS

SCENARIO

	A	B	C	D	I	Total
Slow to detect and respond to COI	6	4	2	0	3	15
Comms w/in team not acted upon	7	5	1	3	5	21
Delay in taking COI under close control	3	3	2	2	2	12
Delay in issuing standard warnings	5	1	0	2	2	10
Premature issuing warnings	1	1	0	0	0	2
Failure to inform higher authority	7	1	1	1	1	11
Lost tactical picture	4	1	4	18	6	33
Failure to use reported sensor info.	2	1	1	1	0	5
Failure to clear clutter	2	1	2	1	1	7
Failure to verify reported track	3	1	1	0	1	6
Failure to take approp responce (ROE)	1	0	2	3	3	9
Failure to use EW softkill	3	1	2	3	2	11
Issued wrong warning level	1	1	0	0	0	2
Failure to ack/act upon Intell msg.	1	1	0	1	0	3
TOTALS:	47	22	18	35	26	148

Table 1. Hot Wash Up Report Results across all teams and all scenarios.

5.3 TapRoot Analysis Data

Detailed examinations of the information processing sequences performed during tactical decision making have revealed a variety of errors. One example involves multiple occurrences, within one scenario, of the substitution of one track number (i.e., a neutral track) for the track number of another (i.e., a declared hostile) aircraft when passing information among team members and when issuing orders to take actions regarding a possible threat aircraft. Other types of errors fall within the following categories: communication, human engineering, procedural, training, and cognitive processing. Examples of communication errors include not using the standard report format for passing information within the team, e.g., not using the track number or bearing when referring to a specific track, and other types of incomplete or incorrect reporting procedures. Examples of human engineering errors include situations where ordered actions were not carried out by team members, confusing track numbers, and confusing surface and air contacts. Procedural errors included not passing information correctly and not ordering a team member to cease issuing warnings to an aircraft who had been identified as a non-threat.

course to remain on a constant heading toward own ship or merely continues on it's original course (no change) would supply a critical piece of evidence. However, only one team from our pilot sessions elected to do this. Would maneuvering own ship have been a very effective and proactive thing to do in this scenario? Yes. Can you score a team negatively for not doing so? No.

This is one of many examples we have come across in our efforts to develop a set of performance measures that we can use in assessing performance data collected in the TADMUS experiments. Similarly, several other actions would be appropriate actions to take to determine the intent of an aircraft. A few teams did illuminate with fire control radar but mostly as a prelude to engaging vice as a means to prod the aircraft pilot into revealing his intent. The employment of decoys is, again, something the ship could do to help assure the pilot sees the ship in the event the pilot was not aware of the ship's presence and may then realize all those warnings he heard during the last few minutes were intended for him. Firing a warning shot (against a surface contact) or a flare (against an aircraft) is another action which might be very effective in eliciting a response from the aircraft--again, something to try--short of engaging. The point to be made is the ship is supposed to take all actions possible to determine whether the inbound aircraft has hostile intent before engaging it. Most teams we've observed to date have not employed several of these actions. A method for assessing decisionmakers on their performance, which is accepted by the operational community, needs to be developed.

6.2 DEFTT Laboratory Acceptability

From feedback collected on the overall fidelity of the DEFTT Lab we learned the following three changes to our procedure would be advantageous to conducting the DEFTT experiment: (1) provide scenario package prior to experiment; (2) train to criteria on DEFTT system use; and (3) use confederates to fill the four support positions.

Providing a scenario package to the subjects (CO and TAO) one week prior to the experimental session has two advantages: (1) it is a more realistic way of exposing the subjects to the pre-experimental material, i.e., the rules of engagement, political/military situation, threat summary, and the ship's mission statement and (2) it saves time during the experimental session. Providing this scenario package before the experiment allows the decisionmaker to integrate the contents of the pre-experimental materials and to think about how he might respond to various threats and situations that he might be confronted with in that part of the world. It makes the task more realistic in that it more closely parallels what the team would actually do aboard ship. The team would normally have several weeks of transiting time to get to their assigned operational station and during these few weeks they would be constantly reviewing the threat in that area, the pertinent rules of engagement, possible scenarios, and so on. A second, less important reason for providing a pregame package is that it saves time by letting subjects review it before arriving in the DEFTT Lab.

When we examined the data from each experimental session we saw indications of a learning effect across the scenarios. Providing additional training on the use of the individual computer workstations should obviate this effect. A criteria-based training package for all subjects is being developed to assure they are familiar with the necessary keystrokes prior to engaging in the experimental test scenarios. This additional training should also serve to lesson a few types of errors that we categorized as resulting from the artificiality of the DEFTT Lab.

From observations during the pilot runs we have learned our main interest is the key decision-makers--the CO and TAO--and their decision-making process. The DSS will directly support only the CO/TAO dyad which means we need to control the influence of the four support team members like any other moderator variable. One way to control this moderator variable is to use confederates. As we observed in the pilot runs the support team affects the decisionmaking process of the CO/TAO dyad, for example, by being late to pass information or by not using required standard communication procedures. We have documented examples from the pilot runs of how support teams negatively influenced experienced COs and TAOs. Moreover, we observed a large degree of variance in the performance of the various support teams. For experimental rigor we decided the optimal way of controlling this variable is to use confederates. This will make the data within and between teams comparable and controllable. A side benefit from the use of confederates is it will enable us to train our own confederate support team members to a certain level of performance which will decrease the time required to train the team during the experimental session.

Failure to clear clutter. The DSS will filter the tracks and only display tracks which are potential COIs. The decisionmaker will be presented with a less cluttered picture on the DSS display compared to the geographic plot on the Aegis Display System where all contacts, including background contacts, are displayed.

Failure to take appropriate response. All the various actions the team is supposed to take will be preprogrammed into the DSS and the user will be prompted to take these actions at the optimal times.

Failure to use softkill. This error should be addressed if the user specifies in advance that he wants to receive a prompt to use softkill tactics. Prompts to fire chaff or launch decoys could be preprogrammed as part of the timeline and its associated alerts.

Issued wrong warning level. This error category has two causes based on our interpretation of team dynamics. In some cases the wrong warning level was issued because the TIC apparently misheard or misinterpreted what the TAO told him. We interpret this case as resulting from the high workload inherent in the scenarios. This miscommunication error would not be directly addressable directly by the DSS. However, we envision the team, in general, will be able to better process the COIs and so, indirectly, the DSS could help decrease the occurrence of this error. The second cause for issuing the wrong warning level was the level of warning to be issued was not specified by the officer who ordered that a warning be given. The DSS will specify to the user which warning level is recommended, thus, the occurrence of this error should be decreased.

Failure to acknowledge or act upon intelligence reports. This type of tactical data, specifically an intelligence message, will automatically be included in the summarized information presented by the DSS

References

Barnett, B. J., Perrin, B. M., & Walrath, L.C. (1993) *Bias in Human Decision Making for Tactical Decision Making Under Stress* (Task 1). Naval Command, Control, and Ocean Surveillance Center, RDT&E Division, San Diego, CA. In press.

Barnett, B. J., Perrin, B. M., & Walrath, L.C. (1993) *Techniques to Reduce Bias in Human Decision Making* for *Tactical Decision Making Under Stress* (Tasks 2 & 3). Naval Command, Control, and Ocean Surveillance Center, RDT&E Division, San Diego, CA. In press.

Hair, D. C., & Pickslay, K. (1992) *Explanation Based Reasoning in Decision Support Systems*. Proceedings of the 9th Annual Conference on Command and Control Decision Aids, June 1992, Monterey, CA.

Hair, D. C., Pickslay, K. & Chow, S. (1992) *Explanation-Based Decision Support in Real Time Situations.* Proceedings of the 1992 IEEE International Conference on Tools with AI. Nov. 1992, Arlington, VA.

Hutchins, S. G. & Duffy, L. T. (1992) *Decision-Making Evaluation Facility for Tactical Teams.* Proceedings of the 9th Annual Conference on Command and Control Decision Aids, June 1992, Monterey, CA.

Kahneman, D., Slovic, P., & Tversky, A. (1982) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press.

Klein, G. A. (1993) A Recognition-Primed Decision (RPD) Model of Rapid Decision Making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.) *Decision Making in Action: Models and Methods.* Ablex Publishing Corporation, New Jersey.

Kramer, T. (1991) *Aegis Lessons Learned.* Presentation to the TADMUS Technical Advisory Board. Naval Command, Control, and Ocean Surveillance Center, RDT&E Division, November 1991, San Diego, CA.

Orasanu, J. & Connolly, T. (1993) The reinvention of Decision Making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.) *Decision Making in Action: Models and Methods.* Ablex Publishing Corporation, New Jersey.

Paradies, M. (1991) Root Cause Analysis and Human Factors. *Human Factors Society Bulletin, 34*(8), 1-4.

Paradies, M. & Unger, L. (1991) *TapRoot Incident Investigation System Manual.* Volumes 1-7. System Improvements, Inc. Knoxville, TN.